The hyperbolic meaning of the Milnor–Wood inequality

Daniel V. Mathews

Abstract

We introduce a notion of the twist of an isometry of the hyperbolic plane. This twist function is defined on the universal covering group of orientation-preserving isometries of the hyperbolic plane, at each point in the plane. We relate this function to a function defined by Milnor and generalised by Wood. We deduce various properties of the twist function, and use it to give new proofs of several well-known results, including the Milnor–Wood inequality, using purely hyperbolic-geometric methods. Our methods express inequalities in Milnor's function as equalities, with the deficiency from equality given by an area in the hyperbolic plane. We find that the twist of certain products found in surface group presentations is equal to the area of certain hyperbolic polygons arising as their fundamental domains.

Contents

1   Introduction

1.1   Overview

In his 1957 paper [18], Milnor introduced a function $\Theta : \widetilde{GL_2^+\mathbb{R}} \longrightarrow \mathbb{R}$ which is in a sense a "rotation angle" associated to elements of the universal covering group of the matrix group $GL_2^+\mathbb{R}$. He proved that it satisfies the inequality

$$\left| \Theta(\alpha\beta) - \Theta(\alpha) - \Theta(\beta) \right| < \frac{\pi}{2},$$

i.e. is a *quasimorphism*, and used it to prove a theorem regarding the existence of principal $GL_2^+\mathbb{R}$ bundles over a closed oriented surface with a flat connection. This result was extended by Wood in [20], who defined a function $r : \widetilde{\mathrm{Top}^+ S^1} \to \mathbb{R}$, with similar properties; here $\mathrm{Top}^+ S^1$ is the group of orientation-preserving homeomorphisms of the circle and $\widetilde{\mathrm{Top}^+ S^1}$ its universal cover. Wood used this function r to prove, *inter alia*, a theorem regarding bundles over surfaces with structure group $\mathrm{Top}\, S^1$; in particular, when the structure group reduces to a totally disconnected subgroup.

One way to interpret the proofs of these theorems, broadly, is as follows. The function Θ or r gives a measure of how far an element of $\tilde{G}$ (where G is $GL_2^+\mathbb{R}$ or $\mathrm{Top}^+ S^1$ or some other group) is from the origin. The quasimorphism property is used to show that a commutator of any two elements in $\tilde{G}$ cannot be "too far" from the origin. Since bundles over surfaces with flat connections (or totally disconnected structure group) are given by holonomy representations, understanding bundles of the desired type is essentially the same as understanding holonomy representations; and since an oriented surface has a standard presentation with one relator, namely a product of commutators, the understanding of commutators in $\tilde{G}$ gives results about the existence of such bundles.

The key result in these theorems, then, is what has become known as the *Milnor–Wood inequality* (see e.g. [10]), which expresses how far a product of commutators in $\tilde{G}$, which multiplies to $1 \in G$ (as required of a surface group representation) can stray from the identity. In particular, letting the lifts of $1 \in G$ to $\tilde{G}$ be $\{\mathbf{z}^m\}$, such a product of commutators is of the form $\mathbf{z}^m$; this m is essentially the *Euler class* of the representation and the content of the inequality is that this Euler class cannot exceed the Euler characteristic of the surface in magnitude.

The Milnor–Wood inequality is by now a classical result and has given rise to a vast array of applications and generalisations. For example: in the theory of Lorentz spacetimes of constant curvature [17], circular groups [6], foliations [19, 7], contact geometry [8], and bounded cohomology [12]. It has been generalised to other Lie groups [2] and to general representations of lattices into Lie groups of Hermitian type [4]. Analogous results exist in higher-dimensional hyperbolic geometry [1] and for manifolds locally isometric to a product of hyperbolic planes [3]. This is just a random sample and is by no means even an overview of the work which exists on the topic.

In this paper we present something far lower-powered, and restricted to Milnor's original case, but perhaps still of interest; we are surprised not to have found this idea in the existing literature. The present paper is concerned with $G = SL_2\mathbb{R}$; obviously $SL_2\mathbb{R} \subset GL_2^+\mathbb{R}$ and $\widetilde{SL_2\mathbb{R}} = \widetilde{PSL_2\mathbb{R}} = \widetilde{\mathrm{Isom}^+ \mathbb{H}^2}$. Milnor's Θ thus assins a number to a (lift to universal cover of a) hyperbolic isometry. We will give a hyperbolic-geometric interpretation of Θ by defining a function

$$\mathrm{Tw} \ : \ \widetilde{PSL_2\mathbb{R}} \times \overline{\mathbb{H}}^2 \longrightarrow \mathbb{R},$$

the "twist angle" of an $\tilde{\alpha} \in \widetilde{PSL_2\mathbb{R}}$ at a point $p \in \mathbb{H}^2$, which generalises Θ. This function has interesting properties, including quasimorphism-type properties, which give a hyperbolic-geometric proof of the quasimorphism property of Θ. Even better, we give an *equality* in which the defect of Tw (and hence Θ) from being a homomorphism is expressed as an area in the hyperbolic plane. Areas arise as deficiencies from additivity essentially because of the effect of negative curvature on parallel translation. Thus, we obtain a new proof of the Milnor–Wood inequality by pure hyperbolic-geometric methods.

We have several other applications. We use the function Tw to prove various relationships between surface group representations and areas in the hyperbolic plane. We interpret the twist of a commutator as the area of a hyperbolic pentagon, and indeed we can interpret the twist of any product occurring as a standard orientable surface group relator as an area of a polygon in the hyperbolic plane. We can also reprove some known results about hyperbolic isometries: which elements of $\widetilde{PSL_2}\mathbb{R}$ which can occur as commutators [20, 7, 10]; relationships between types of commutators and their trace [10, 11]; and a cute result, as far as we know first appearing in [11], characterising isometries of hyperbolic type with intersecting axes in terms of the trace of their commutator.

However, in our view the main new application of these methods is in a pair of subsequent papers, where we consider the question of which representations of the fundamental group of a surface are holonomy representations of hyperbolic structures, and cone-manifold structures of a certain type: see [14, 15]. Our methods here establish connections between the algebra of $PSL_2\mathbb{R}$ and hyperbolic geometry, which we use in those papers.

We finally note that $SL_2\mathbb{R} \cong \mathrm{Sp}\,(2)$, the group of 2×2 symplectic matrices, i.e. linear symplectomorphisms of $\mathbb{R}^2$ with the standard symplectic structure. Milnor's function Θ in this context is essentially the *Maslov index* (see e.g. [16, p. 48]). We wonder if there are any further connections to symplectic geometry.

1.2   Structure of this paper

In section 2 we define the notion of twist. This first requires some preliminaries on $\widetilde{PSL_2}\mathbb{R}$, which occupy sections 2.1 to 2.3. In section 3 we establish various properties of our twist function. In section 4 we recall the definition of Milnor's function Θ, we relate it to twisting, and deduce various properties.

1.3   Acknowledgments

2   Twisting in the hyperbolic plane

Everything in sections 2.1 to 2.3 has been known for a long time: see, e.g. [10]. Although the idea is very basic, it appears that the notion of the twist function which we define in section 2.4 is new.

2.1 $PSL_2\mathbb{R}$ and $\widetilde{PSL_2}\mathbb{R}$

Fix a basepoint y_0 in $\mathbb{H}^2$ and unit tangent vector $u_0 \in UT_{y_0}\mathbb{H}^2$. An orientation-preserving hyperbolic isometry is uniquely determined by the image of u_0, and we may identify the unit tangent bundle $UT\mathbb{H}^2$ with the orientation-preserving isometry group $PSL_2\mathbb{R}$. Topologically $PSL_2\mathbb{R} \cong \mathbb{R}^2 \times S^1$; let p_1 be the projection map $PSL_2\mathbb{R} \longrightarrow \mathbb{H}^2$.

Let $p_2 : \widetilde{PSL_2}\mathbb{R} \longrightarrow PSL_2\mathbb{R}$ be the universal cover of $PSL_2\mathbb{R}$; see [9, 10] for further details. Clearly $\pi_1(PSL_2\mathbb{R}) \cong \mathbb{Z}$. An element $\tilde{x} \in \widetilde{PSL_2}\mathbb{R}$ is hyperbolic, elliptic or parabolic accordingly as is $p_2(\tilde{x}) \in PSL_2\mathbb{R}$. We can consider $\widetilde{PSL_2}\mathbb{R}$ as $\mathbb{H}^2$, with an $\mathbb{R}$ fibre above each point, covering the circle of unit tangent vectors.

$$\widetilde{PSL_2}\mathbb{R} \xrightarrow{p_2} PSL_2\mathbb{R} \cong UT\mathbb{H}^2 \xrightarrow{p_1} \mathbb{H}^2.$$

We can also consider elements of $\widetilde{PSL_2}\mathbb{R}$ as homotopy classes of paths in $UT\mathbb{H}^2$ starting at the basepoint. Since the basepoint is arbitrary, *every* path $c : [0,1] \longrightarrow UT\mathbb{H}^2$ determines a unique

element of $\widetilde{PSL_2\mathbb{R}}$, which we also denote c, abusing notation. The projection of c to $PSL_2\mathbb{R}$ is the orientation-preserving isometry sending $c(0)$ to $c(1)$. An $\alpha \in PSL_2\mathbb{R}$ has countably infinitely many lifts to $\widetilde{PSL_2\mathbb{R}}$. These all represent paths in $UT\mathbb{H}^2$ between the same start and end tangent vectors. However these paths will differ according to the number of times that the tangent vectors spin as the path is traversed. The lifts of the identity $1 \in PSL_2\mathbb{R}$ form an infinite cyclic group $\{\mathbf{z}^n : n \in \mathbb{Z}\}$, where $\mathbf{z}$ is the homotopy class of the path $c(t) = (y_0, e^{2\pi i n t} u_0)$. Note $\mathbf{z}$ commutes with every element of $\widetilde{PSL_2\mathbb{R}}$; in fact $\mathbf{z}$ generates the centre of $\widetilde{PSL_2\mathbb{R}}$.

2.2 Regions in $\widetilde{PSL_2\mathbb{R}}$

While every element has infinitely many lifts, some lifts are simpler than others. For instance, the identity in $\widetilde{PSL_2\mathbb{R}}$ is the "simplest" lift of the identity in $PSL_2\mathbb{R}$.

If $\alpha \in PSL_2\mathbb{R}$ is hyperbolic then it translates by distance d_α along Axis α. Let $c(t) \in PSL_2\mathbb{R}$ be the translation of (signed) hyperbolic distance td_α along Axis α; then $c : \mathbb{R} \longrightarrow PSL_2\mathbb{R}$ is a homomorphism with $c(1) = \alpha$, in fact the only homomorphism with this property. The path $c|_{[0,1]}$ in $PSL_2\mathbb{R}$ gives an element $\tilde{\alpha}$ of $\widetilde{PSL_2\mathbb{R}}$ which we take as our preferred or simplest lift. This lift can be considered a path of unit tangent vectors, which travels along Axis α at speed d, always pointing along Axis α in the direction of translation.
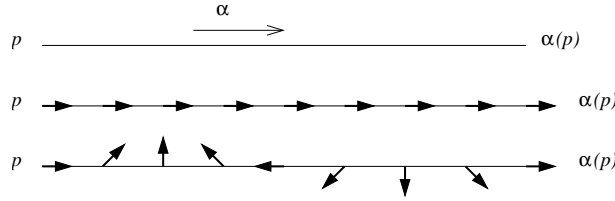

Figure 1: An isometry $\alpha \in PSL_2\mathbb{R}$; the simplest lift of α; a different lift of α.
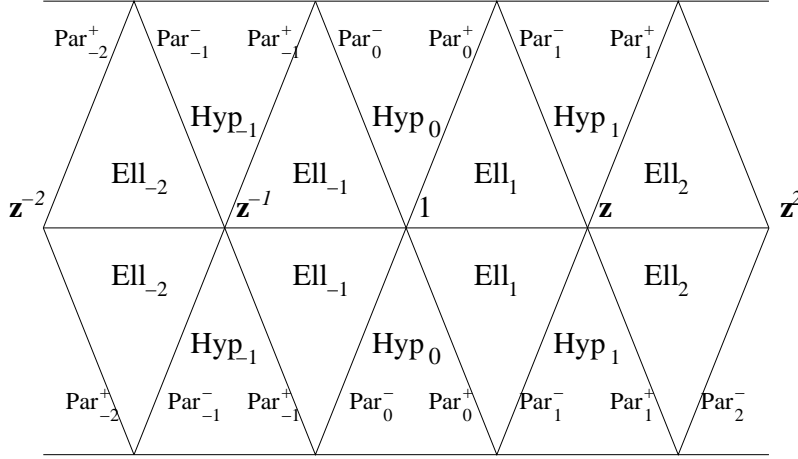
Similar considerations apply to parabolic isometries. A parabolic $\alpha \in PSL_2\mathbb{R}$ translates along some horocycle h_α (not unique); endowing h_α with a Euclidean metric, let α translate by Euclidean distance d. Letting $c(t) \in PSL_2\mathbb{R}$ be the parabolic translating td along h_α then $c : \mathbb{R} \longrightarrow PSL_2\mathbb{R}$ is the unique homomorphism with $c(1) = \alpha$, and $c|_{[0,1]}$ gives a preferred lift $\tilde{\alpha} \in \widetilde{PSL_2\mathbb{R}}$. This $\tilde{\alpha}$ can be considered a path of tangent vectors travelling along and pointing along h_α at speed d for time 1.

However the situation for elliptic $\alpha \in PSL_2\mathbb{R}$ is different: there are infinitely many homomorphisms $c : \mathbb{R} \longrightarrow PSL_2\mathbb{R}$ with $c(1) = \alpha$. Let α rotate by angle θ (mod 2π). Then the lifts of α are rotations by angles $\theta + 2\pi\mathbb{Z}$. From this viewpoint there are two simplest lifts of α, those with rotation angle lying in $(0, 2\pi)$ and $(-2\pi, 0)$: a simplest anticlockwise and clockwise lift.

Denote the sets of simplest lifts of hyperbolics and parabolics Hyp_0 and Par_0 respectively. Let $\mathrm{Hyp}_n = \mathbf{z}^n \mathrm{Hyp}_0$ and $\mathrm{Par}_n = \mathbf{z}^n \mathrm{Par}_0$, so the hyperbolic (resp. parabolic) elements of $\widetilde{PSL_2\mathbb{R}}$ are $\sqcup_n \mathrm{Hyp}_n$ (resp. $\sqcup_n \mathrm{Par}_n$). We may consider $\tilde{\alpha} \in \mathrm{Hyp}_n$ as a translation along Axis α with an added twist of $2n\pi$. We may further distinguish between Par_n^+ and Par_n^-, the rotations about points at infinity whose projections to $PSL_2\mathbb{R}$ are anticlockwise and clockwise respectively.

As for elliptics, let the set of simplest anticlockwise and clockwise lifts be Ell_1 and Ell_{-1} respectively. For $n > 0$ let $\mathrm{Ell}_n = \mathbf{z}^{n-1} \mathrm{Ell}_1$ and $\mathrm{Ell}_{-n} = \mathbf{z}^{-n+1} \mathrm{Ell}_{-1}$. (So Ell_0 is not defined and $\mathbf{z} \mathrm{Ell}_{-1} = \mathrm{Ell}_1$.) For $n > 0$ (resp. $n < 0$), Ell_n consists of all rotations through angles between $2\pi(n-1)$ and $2\pi n$ (resp. between $2\pi n$ and $2\pi(n+1)$).

Considering that hyperbolics and elliptics form 3-dimensional subspaces of the 3-dimensional $PSL_2\mathbb{R}$, with common 2-dimensional boundary the space of parabolic elements, we may draw a schematic diagram of $\widetilde{PSL_2\mathbb{R}}$ as in figure 2.

Figure 2: Schematic diagram of $\widetilde{PSL_2}\mathbb{R}$.

Lemma 2.1 *Let $\alpha, \beta \in PSL_2\mathbb{R}$. Then $[\alpha, \beta]$ has a well-defined lift to $\widetilde{PSL_2}\mathbb{R}$. That is, any two sets of lifts $\tilde{\alpha}_1, \tilde{\beta}_1$ and $\tilde{\alpha}_2, \tilde{\beta}_2$ satisfy $[\tilde{\alpha}_1, \tilde{\beta}_1] = [\tilde{\alpha}_2, \tilde{\beta}_2]$.*

PROOF Let $\tilde{\alpha}_2 = \mathbf{z}^a \tilde{\alpha}_1$, $\tilde{\beta}_2 = \mathbf{z}^b \tilde{\beta}_1$. Since $\mathbf{z}$ is central,

$$[\tilde{\alpha}_2, \tilde{\beta}_2] = \tilde{\alpha}_2 \tilde{\beta}_2 \tilde{\alpha}_2{}^{-1} \tilde{\beta}_2{}^{-1} = \mathbf{z}^a \tilde{\alpha}_1 \mathbf{z}^b \tilde{\beta}_1 \tilde{\alpha}_1{}^{-1} \mathbf{z}^{-a} \tilde{\beta}_1{}^{-1} \mathbf{z}^{-b} = \tilde{\alpha}_1 \tilde{\beta}_1 \tilde{\alpha}_1{}^{-1} \tilde{\beta}_1{}^{-1} = [\tilde{\alpha}_1, \tilde{\beta}_1].$$

■

2.3 Traces in $\widetilde{PSL_2}\mathbb{R}$

As $\widetilde{PSL_2}\mathbb{R}$ covers $SL_2\mathbb{R}$, there is a well-defined trace on $\widetilde{PSL_2}\mathbb{R}$. For all elliptic regions, the trace lies in $(-2, 2)$; in the various other regions of $\widetilde{PSL_2}\mathbb{R}$ it takes values as follows.

Lemma 2.2 (Trace lemma)

$$\text{Tr}(\mathbf{z}^n) = (-1)^n \cdot 2, \quad \text{Tr}(\text{Par}_n) = (-1)^n \cdot 2, \quad \text{Tr}(\text{Hyp}_n) = \begin{cases} (2, \infty) & n \; even \\ (-\infty, -2) & n \; odd. \end{cases}$$

PROOF Consider the matrix

$$E(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Now in the upper half plane $E(\theta)$ is elliptic, a rotation of 2θ about i. Thus $E(n\pi) = \mathbf{z}^n$, and hence $\mathbf{z}^n$ projects to $(-1)^n \in SL_2\mathbb{R}$. From this the first claim follows. The trace of an element of $\widetilde{PSL_2}\mathbb{R}$ is ± 2 if and only if it is a power of $\mathbf{z}$, or parabolic. Now Tr is a continuous function and, considering the topology of $\widetilde{PSL_2}\mathbb{R}$, $\text{Par}_n \cup \{\mathbf{z}^n\}$ is connected. Thus $\text{Tr}(\text{Par}_n) = (-1)^n$. As Hyp_n is connected and bounded by Par_n and $\mathbf{z}^n$, on which $\text{Tr} = (-1)^n \cdot 2$, the final claim follows. ■

2.4 Definition of twist

Define the *twist of a vector field along a curve* as follows. Consider a smooth curve $c : [0, 1] \longrightarrow \mathbb{H}^2$ and a smooth unit tangent vector field $\mathcal{V} : [0, 1] \longrightarrow UT\mathbb{H}^2$, $p_1 \circ \mathcal{V} = c$ (recall p_1 is the projection $UT\mathbb{H}^2 \longrightarrow \mathbb{H}^2$). Consider the velocity vector field $\frac{dc}{dt}$ along c, which we may rescale to a unit vector field $\hat{c} : [0, 1] \longrightarrow UT\mathbb{H}^2$. Consider the angle $\theta(t)$ (measured anticlockwise) from $\hat{c}(t)$ to $\mathcal{V}(t)$. We
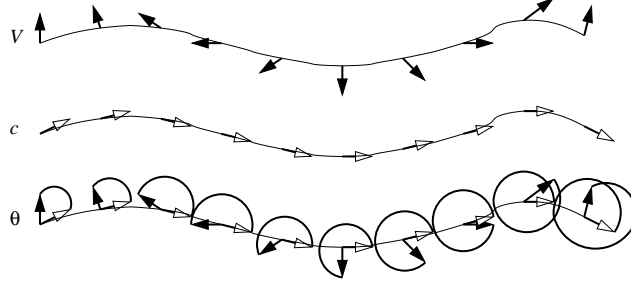
Figure 3: The twist of a vector field along a curve.

have many choices for $\theta(0)$ (differing by $2\pi\mathbb{Z}$), but choosing $\theta(0)$ arbitrarily determines continuous θ completely; $\theta(1) - \theta(0)$ is independent of this choice, and is the twist of $\mathcal{V}$ along c.

Now given $y \in \mathbb{H}^2$ and $\tilde{\alpha} \in \widetilde{PSL_2\mathbb{R}}$ we define the *twist of $\tilde{\alpha}$ at y*, denoted $\mathrm{Tw}(\tilde{\alpha}, y)$. Let $\tilde{\alpha}$ project to $\alpha \in PSL_2\mathbb{R}$. Let $c : [0,1] \longrightarrow \mathbb{H}^2$ be a constant speed (possibly 0) geodesic from y to $\alpha(y)$. There is a vector field $\mathcal{V} : [0,1] \longrightarrow UT\mathbb{H}^2$ along c which lies in the homotopy class of $\tilde{\alpha}$. Then $\mathrm{Tw}(\tilde{\alpha}, y)$ is the twist of $\mathcal{V}$ along c.

That is, $\mathrm{Tw}(\tilde{\alpha}, y)$ describes how the tangent vector at y is moved by $\tilde{\alpha}$, compared to parallel translation along the geodesic from y to $\tilde{\alpha}(y)$. It is clear this does not depend on the choice of $\mathcal{V}$. For $\alpha \in PSL_2\mathbb{R}$, define $\mathrm{Tw}(\alpha, y)$ the same way, except angles are taken modulo 2π.

As an aside, we note that this method of obtaining a rotation angle from an element of $\widetilde{PSL_2\mathbb{R}}$ is not so different from Wood's method in [20]. Wood regards an element of $SL_2\mathbb{R}$ as acting on $\mathbb{R}^2$ by a linear transformation, and hence on the S^1 of oriented lines through the origin. Thus there is an inclusion $SL_2\mathbb{R} \subset \mathrm{Top}^+ S^1$. This is equivalent to the action of a hyperbolic isometry on the circle at infinity. But here we regard the isometry as acting on the S^1 of unit tangent vectors at a given point; such unit tangent vectors of course correspond bijectively with the circle at infinity, but different points give different bijections. Our twist is the action of an isometry on unit tangent $\widetilde{S^1}$'s, where the S^1's at a point and its image are related by parallel translation. Wood's function on $\widetilde{\mathrm{Top}^+ S^1}$ involves an integral and hence the measure on the circle at infinity; an isometry however alters this measure.

3 Properties of twisting

3.1 Types of isometries

One can easily verify the following properties of the twist.

- For hyperbolic $\alpha \in PSL_2\mathbb{R}$ and $y \in \mathrm{Axis}\,\alpha$, $\mathrm{Tw}(\alpha, y) = 0 \pmod{2\pi}$. For $\tilde{\alpha} \in \mathrm{Hyp}_0$, $\mathrm{Tw}(\tilde{\alpha}, y) \in (-\pi, \pi)$. The twist is constant along curves of constant distance h from $\mathrm{Axis}\,\alpha$. For each $\theta \in (-\pi, \pi)$ there is precisely one h for which the curve at distance h is the locus of points y with $\mathrm{Tw}(\tilde{\alpha}, y) = \theta$.

- For $\tilde{\alpha} \in \mathrm{Par}_0$, $\mathrm{Tw}(\tilde{\alpha}, y)$ is constant along horocycles about $\mathrm{Fix}\,\alpha$. If $\tilde{\alpha} \in \mathrm{Par}_0^+$ (resp. Par_0^-) then $\mathrm{Tw}(\tilde{\alpha}, y) \in (0, \pi)$ (resp. $(-\pi, 0)$). For horocycles close to $\mathrm{Fix}\,\alpha$, the twist is close to 0. For each $\theta \in (0, \pi)$ (resp. $(-\pi, 0)$) there is precisely one horocycle which is the locus of points y with $\mathrm{Tw}(\tilde{\alpha}, y) = \theta$.

- For elliptic $\tilde{\alpha}$, $\mathrm{Tw}(\tilde{\alpha}, y)$ is constant along hyperbolic circles centred at $\mathrm{Fix}\,\alpha$. Take $\tilde{\alpha} \in \mathrm{Ell}_1$ for convenience, so $\tilde{\alpha}$ rotates by angle $\psi \in (0, 2\pi)$. So $\mathrm{Tw}(\tilde{\alpha}, \mathrm{Fix}\,\alpha) = \psi$. If $\psi \in (0, \pi)$ then $\mathrm{Tw}(\tilde{\alpha}, y)$ always lies in $[\psi, \pi)$; for each $\theta \in (\psi, \pi)$ there is precisely one hyperbolic circle centred at $\mathrm{Fix}\,\alpha$ with is the locus of y with $\mathrm{Tw}(\tilde{\alpha}, y) = \theta$. If $\psi = \pi$ then α is a half turn and $\mathrm{Tw}(\tilde{\alpha}, y) = \pi$ for all

y. If $\psi \in (\pi, 2\pi)$ then $\mathrm{Tw}(\tilde{\alpha}, y)$ always lies in $(\pi, \psi]$ and for each $\theta \in (\psi, \pi)$ there is precisely one hyperbolic circle centred at $\mathrm{Fix}\,\alpha$ which is the locus of y with $\mathrm{Tw}(\tilde{\alpha}, y) = \theta$.

The values of the twist for all Hyp_n, Par_n and Ell_n follow obviously from the above.

Proposition 3.1

$$\mathrm{Tw}(\mathrm{Hyp}_n, \mathbb{H}^2) = \Big((2n-1)\,\pi, (2n+1)\,\pi \Big)$$

$$\mathrm{Tw}(\mathrm{Par}_n, \mathbb{H}^2) = \Big((2n-1)\,\pi, (2n+1)\,\pi \Big)$$

$$\mathrm{Tw}(\mathrm{Ell}_n, \mathbb{H}^2) = \begin{cases} \Big((2n-2)\pi, 2n\pi \Big) & \text{for } n > 0 \\ \Big(-2|n|\pi, (-2|n|+1)\pi \Big) & \text{for } n < 0 \end{cases}$$

■

3.2   Extension to infinity

We have defined $\mathrm{Tw}\,(\tilde{\alpha}, p)$ for $p \in \mathbb{H}^2$. We can extend the definition to $p \in \overline{\mathbb{H}}^2$, with the circle at infinity S^1_∞ adjoined. However we pay the price that, at least if S^1_∞ is endowed with the usual topology of S^1, then Tw is not continuous on S^1_∞.

Extending the definition is simple enough. Take $\tilde{\alpha} \in \widetilde{PSL_2\mathbb{R}}$ and $p \in S^1_\infty$, We note that for any geodesic l with an endpoint at p, if we take points $x \in l$ approaching p then $\mathrm{Tw}\,(\tilde{\alpha}, x)$ approaches a limit; and the limit is independent of choice of l. In particular:

- For $\tilde{\alpha} \in \mathrm{Hyp}_0$, $\mathrm{Tw}(\tilde{\alpha}, p) \in \{-\pi, 0, \pi\}$. The twist is 0 at the two endpoints of $\mathrm{Axis}\,\alpha$ and $\pm\pi$ on the two open arcs of S^∞_1 to either side.

- For $\tilde{\alpha} \in \mathrm{Par}_0^+$ (resp. Par_0^-), $\mathrm{Tw}(\tilde{\alpha}, p)$ is 0 when $p = \mathrm{Fix}\,\alpha$ and π (resp. $-\pi$) otherwise.

- For $\tilde{\alpha} \in \mathrm{Ell}_1$ (resp. Ell_{-1}) then $\mathrm{Tw}(\tilde{\alpha}, p) = \pi$ (resp. $-\pi$) for all $p \in S^1_\infty$.

For $\tilde{\alpha}$ in Hyp_n, $\mathrm{Par}_n^{\pm}$ or Ell_n in general, we adjust by the appropriate multiple of 2π. In particular, on S^1_∞, the twist is always an integer multiple of π.

3.3   Parallel translation and curvature

Recall that curvature is the effect on tangent vectors induced by parallel translation around a loop. As the hyperbolic plane has constant curvature -1, parallel translation around a loop gives a rotation on a tangent vector equal to the negative area enclosed.

Let ABC be a hyperbolic triangle shown in figure 4; let $\alpha, \beta, \gamma \in \mathrm{Isom}^+ \mathbb{H}^2$ respectively be the hyperbolic translations along axes BC, CA, AB and translating $B \mapsto C \mapsto A \mapsto B$. Let $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} \in \mathrm{Hyp}_0$ be their simplest lifts. Then the composition $\tilde{\gamma} \circ \tilde{\beta} \circ \tilde{\alpha}$ is parallel translation around ABC, hence a rotation of signed angle $\theta_A + \theta_B + \theta_C - \pi = -\Delta$, where Δ is the area of ABC. Hence, parallel translation from B to A is equivalent to first rotating an angle of Δ at B, then parallel translating $B \mapsto C \mapsto A$.

Let now $\tilde{\alpha}, \tilde{\beta}$ be *any* elements of $\widetilde{PSL_2\mathbb{R}}$ covering hyperbolic isometries of $\mathbb{H}^2$ which take $B \mapsto C$ and $C \mapsto A$ respectively. Then $\tilde{\beta}\tilde{\alpha}$ takes $B \mapsto A$ and $\mathrm{Tw}(\tilde{\beta}\tilde{\alpha}, B)$ is given by the twist of $\tilde{\beta}\tilde{\alpha}$ along BA relative to parallel translation. If we instead measured the twist of $\tilde{\beta}\tilde{\alpha}$ relative to parallel translation along $B \mapsto C \mapsto A$, i.e. $\mathrm{Tw}(\tilde{\alpha}, B) + \mathrm{Tw}(\tilde{\beta}, C)$, the answer must differ by Δ. Hence we have the following result. Here and below we write $\Delta[A, B, C]$ to denote the signed area of the triangle with ordered vertices A, B, C, and use Δ in general to signify area. Taking a limit of points going to infinity, the result also holds for ideal points.
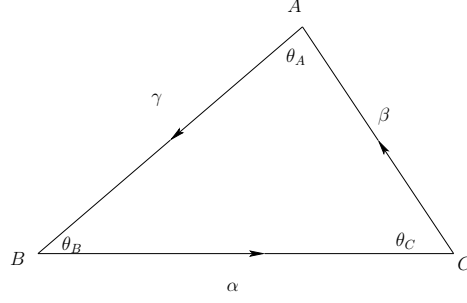
Figure 4: Translation around a hyperbolic triangle.

Lemma 3.2 (Composition lemma) *For any $\tilde{\alpha}, \tilde{\beta} \in \widetilde{PSL_2\mathbb{R}}$ and any $p \in \overline{\mathbb{H}}^2$,*

$$\mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) = \mathrm{Tw}\left(\tilde{\alpha}, p\right) + \mathrm{Tw}\left(\tilde{\beta}, \alpha p\right) - \Delta[p, \alpha p, \beta\alpha p],$$

where $\alpha, \beta \in \mathrm{Isom}^+ \mathbb{H}^2$ are the images of $\tilde{\alpha}, \tilde{\beta}$. ∎

Thus, the failure of Tw to be linear is a manifestation of negative curvature, and the defect is the area of the triangle around which vectors are translated. The defect is clearly bounded as hyperbolic triangles have area less than π. For $p \in \mathbb{H}^2$ then,

$$\left| \mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) - \mathrm{Tw}\left(\tilde{\alpha}, p\right) - \mathrm{Tw}\left(\tilde{\beta}, \alpha p\right) \right| < \pi.$$

For $p \in S_\infty^1$ the inequality holds but is not strict.

3.4 Conjugation and addition

Lemma 3.3 (Conjugation lemma) *For any $\tilde{\alpha}, \tilde{\beta} \in \widetilde{PSL_2\mathbb{R}}$ and $p \in \overline{\mathbb{H}}^2$, $\mathrm{Tw}(\tilde{\alpha}, p) = \mathrm{Tw}(\tilde{\beta}\tilde{\alpha}\tilde{\beta}^{-1}, \beta p)$.*

PROOF Consider $\mathrm{Tw}(\tilde{\alpha}, p)$, i.e. $\tilde{\alpha}$ as a path of unit tangent vectors along the geodesic from p to αp, compared to parallel translation. Now translate the whole situation by the isometry β. ∎

When $\tilde{\beta} = \tilde{\alpha}^n$, for $n \in \mathbb{Z}$, this becomes

$$\mathrm{Tw}\left(\tilde{\alpha}, p\right) = \mathrm{Tw}\left(\tilde{\alpha}, \alpha^n p\right), \tag{3.1}$$

which is clear geometrically: $\alpha^n p$ lies on the same constant distance curve from $\mathrm{Fix}\,\alpha$ or $\mathrm{Axis}\,\alpha$ as p.

Lemma 3.4 (Inverse lemma) *For any $\tilde{\alpha} \in \widetilde{PSL_2\mathbb{R}}$ and $p \in \overline{\mathbb{H}}^2$,*

$$\mathrm{Tw}\left(\tilde{\alpha}, p\right) = -\,\mathrm{Tw}\left(\tilde{\alpha}^{-1}, p\right)$$

PROOF Reversing the path of unit tangent vectors of $\tilde{\alpha}$ gives immediately $\mathrm{Tw}(\tilde{\alpha}, p) = -\mathrm{Tw}(\tilde{\alpha}^{-1}, \alpha p)$. Now apply the previous corollary. ∎

Apply composition lemma 3.2 to the product $\tilde{\beta}\tilde{\alpha} = \left(\tilde{\beta}\tilde{\alpha}\tilde{\beta}^{-1}\right)\tilde{\beta}$

$$\mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) = \mathrm{Tw}\left(\tilde{\beta}, p\right) + \mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}\tilde{\beta}^{-1}, \beta p\right) - \Delta[p, \beta p, \beta\alpha p],$$

then apply conjugation lemma 3.3. This gives a result like 3.2, but now all based at the same $p \in \mathbb{H}^2$.

Lemma 3.5 (Addition lemma) *For all* $\tilde{\alpha}, \tilde{\beta} \in \widetilde{PSL_2\mathbb{R}}$ *and* $p \in \overline{\mathbb{H}}^2$,

$$\mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) = \mathrm{Tw}\left(\tilde{\beta}, p\right) + \mathrm{Tw}\left(\tilde{\alpha}, p\right) - \Delta[p, \beta p, \beta \alpha p].$$ ∎

This "addition formula" for Tw describes the twist of a product in terms of the twist of the factors, all at the same point. We immediately obtain a quasimorphism property of Tw: for $p \in \mathbb{H}^2$,

$$\left|\mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) - \mathrm{Tw}\left(\tilde{\beta}, p\right) - \mathrm{Tw}\left(\tilde{\alpha}, p\right)\right| < \pi. \tag{3.2}$$

If $p \in S_1^\infty$ the inequality is not strict.

3.5   Pentagons and commutators

Definition 3.6 *Let* $\alpha, \beta \in PSL_2\mathbb{R}$ *and* $p \in \mathbb{H}^2$. *Then the geodesic pentagon in* $\mathbb{H}^2$ *obtained by joining the segments*

$$p \longrightarrow \alpha^{-1}\beta^{-1}\alpha\beta p \longrightarrow \beta p \longrightarrow \alpha\beta p \longrightarrow \beta^{-1}\alpha\beta p \longrightarrow p$$

is called the pentagon generated by α, β at p *and is denoted* $\mathcal{P}(\alpha, \beta; p)$.

Note that $\mathcal{P}(\alpha, \beta; p)$ may intersect itself; its vertices may coincide; it need not even bound an immersed disc. It is simply five geodesic line segments, possibly degenerate, in $\mathbb{H}^2$; if these all have nonzero length we say $\mathcal{P}(\alpha, \beta; p)$ is *nondegenerate*. Denote the vertices

$$p_0 = p, \quad p_1 = \beta p, \quad p_2 = \alpha\beta p, \quad p_3 = \beta^{-1}\alpha\beta p, \quad p_4 = \alpha^{-1}\beta^{-1}\alpha\beta p.$$

If $\mathcal{P}(\alpha, \beta; p)$ is nondegenerate and bounds an immersed disc, denote the interior angles of the pentagon $\theta_0, \ldots, \theta_4$. We may also denote a polygon as the sequence of vertices; we write $\mathcal{P}(\alpha, \beta; p) = [p_0, p_4, p_1, p_2, p_3]$.

If $\mathcal{P}(\alpha, \beta; p)$ bounds an embedded disc, then it has a well-defined area $\Delta[\mathcal{P}(\alpha, \beta; p)]$. This area is signed according to the boundary orientation $p_0 \to p_4 \to p_1 \to p_2 \to p_3 \to p_0$. The same can be done even if $\mathcal{P}(\alpha, \beta; p)$ bounds an immersed disc (see figure 5); for instance by cutting into smaller embedded discs.
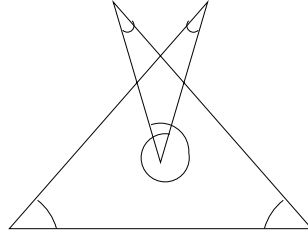

Figure 5: $\mathcal{P}$ may bound an immersed but not embedded disc.

Proposition 3.7 (Commutator pentagon) *If* $\mathcal{P}(\alpha, \beta; p)$ *is nondegenerate and bounds an immersed disc, then*

$$\mathrm{Tw}\left([\tilde{\alpha}^{-1}, \tilde{\beta}^{-1}], p\right) = \Delta[\mathcal{P}(\alpha, \beta; p)].$$

(Recall by lemma 2.1, the commutator is independent of choice of lifts $\tilde{\alpha}, \tilde{\beta}$.)

PROOF Without loss of generality assume $\Delta > 0$. Consider figure 6. Consider a unit tangent vector (p, u) at p pointing along the geodesic to p_4. Follow ("chase") the image of this vector under $D\beta$, $D\alpha$, $D\beta^{-1}$ and $D\alpha^{-1}$ to obtain unit tangent vectors (p_i, u_i) at each p_i. Note that α takes the segment $p_4 \to p_1$ to the segment $p_3 \to p_2$ and β takes $p_0 \to p_3$ to $p_1 \to p_2$; we use these two facts repeatedly.
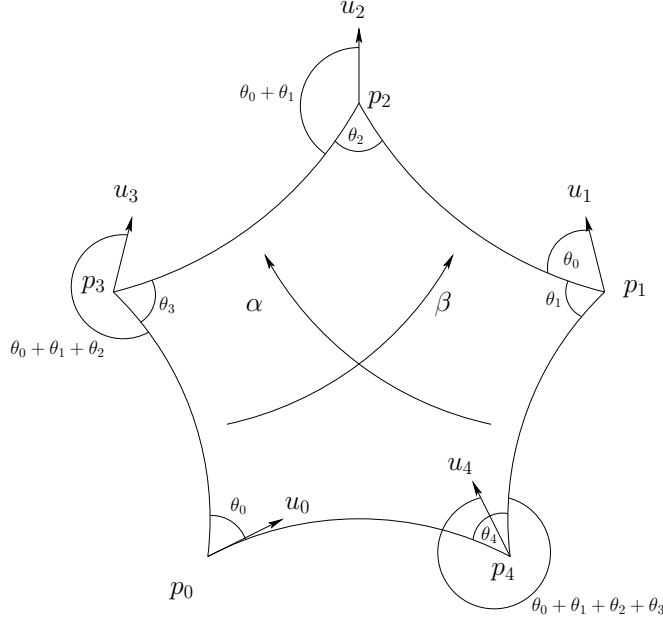
Figure 6: A unit vector chase.

Now $(p_0, u_0) = (p, u)$ is based at p and points θ_0 clockwise of $p_0 \to p_3$; hence $(p_1, u_1) = D\beta(p, u)$ is based at p_1 and points θ_0 clockwise of $p_1 \to p_2$. But (p_1, u_1) points $\theta_0 + \theta_1$ clockwise of $p_1 \to p_4$, hence $(p_2, u_2) = D\alpha(p_1, u_1)$ points $\theta_0 + \theta_1$ clockwise of $p_2 \to p_3$. Then (p_2, u_2) lies $\theta_0 + \theta_1 + \theta_2$ clockwise of $p_2 \to p_1$, hence $(p_3, u_3) = D\beta^{-1}(p_2, u_2)$ lies the same angle clockwise of $p_3 \to p_0$. Finally, (p_3, u_3) lies $\theta_0 + \theta_1 + \theta_2 + \theta_3$ clockwise of $p_3 \to p_2$, and hence $(p_4, u_4) = D[\alpha^{-1}, \beta^{-1}](p, u)$ lies the same angle clockwise of $p_4 \to p_1$. Thus $D[\alpha^{-1}, \beta^{-1}](p, u)$ lies $\theta_0 + \theta_1 + \theta_2 + \theta_3 + \theta_4$ clockwise of $p_4 \to p_0$.

This immediately shows that $\mathrm{Tw}([\alpha^{-1}, \beta^{-1}], p) = 3\pi - \sum \theta_i = \Delta[\mathcal{P}(\alpha, \beta; p)]$, modulo 2π. Choosing particular lifts $\tilde{\alpha}, \tilde{\beta}$ of α, β we may see that this is true over the real numbers. For this we use the following lemma, which is straightforward, although in the elliptic case perhaps the reader might draw a few pictures to convince herself.

Lemma 3.8 *Let $\alpha \in PSL_2\mathbb{R}$ be an isometry and $q_1 \neq q_2$ be two points in $\mathbb{H}^2$, neither of which is fixed by α. Let l denote the geodesic from q_1 to q_2, l_1 the geodesic segment from q_1 to αq_1 and l_2 the geodesic segment from q_2 to αq_2. Suppose we have two vector fields $\mathcal{V}_1, \mathcal{V}_2$ along l_1, l_2 respectively, and $c \in \mathbb{Z}$ such that:*

(i) $\mathcal{V}_1$ begins at q_1 pointing along l towards q_2; ends at αq_1 pointing along αl towards αq_2; crosses the direction of l_1 transversely at finitely many points; the crossings taken with sign sum to c;

(ii) $\mathcal{V}_2$ begins at q_2 pointing along l towards q_1; ends at αq_2 pointing along αl towards αq_1; crosses the direction of l_2 transversely at finitely many points; the crossings taken with sign sum to c.

Then the vector fields $\mathcal{V}_1, \mathcal{V}_2$ both represent the same lift of α to $\widetilde{PSL_2}\mathbb{R}$. ∎

Since β takes the segment $p_0 \to p_3$ to $p_1 \to p_2$, we may choose $\tilde{\beta}$ to be represented by a path of unit tangent vectors along the geodesic $p_0 \to p_1$ which begins pointing along $p_0 \to p_3$ and ends pointing along $p_1 \to p_2$; we may also choose $\tilde{\beta}$ to be represented by a path of unit tangent vectors along $p_3 \to p_2$, which begins pointing along $p_3 \to p_0$ and ends pointing along $p_2 \to p_1$. Using the above lemma, we may ensure that these two paths of tangent vectors represent the same $\tilde{\beta}$. Similarly, since α takes $p_4 \to p_1$ to $p_3 \to p_2$ we may choose $\tilde{\alpha}$ to be represented by the path of unit tangent vectors

along the geodesic $p_4 \to p_3$ which begins pointing along $p_4 \to p_1$, ends pointing along $p_3 \to p_2$; and also we may choose $\tilde{\alpha}$ to be represented by tangent vectors along $p_1 \to p_2$ which begins pointing along $p_1 \to p_4$ and ends pointing along $p_2 \to p_3$. Chasing these paths of vectors around the pentagon, then, we obtain $\mathrm{Tw}([\tilde{\alpha}^{-1}, \tilde{\beta}^{-1}], p) = \Delta[\mathcal{P}(\alpha, \beta; p)]$ on the nose. ∎

It follows from the above proof that, even if $\mathcal{P}(\alpha, \beta; p)$ does not bound an immersed disc, we may still follow unit vectors and obtain $\mathrm{Tw}([\alpha^{-1}, \beta^{-1}], p) \equiv 3\pi - \sum \theta_i$ modulo 2π, where θ_i are the various angles between segments of $\mathcal{P}(\alpha, \beta; p)$.

If S is a punctured torus with a hyperbolic structure and totally geodesic boundary, then a pentagon $\mathcal{P}(\alpha, \beta; p)$ is a fundamental domain for S, where α, β are the holonomy of a meridian and longitude, for appropriate choice of p, and $\mathcal{P}(\alpha, \beta; p)$ bounds an embedded disc; thus the area $\Delta[S]$ of S is the area of the pentagon. By Gauss-Bonnet this area is 2π. Also $[\alpha, \beta]$ is the holonomy of the boundary curve. Choosing $p \in \mathbb{H}^2$ as the appropriate vertex of the fundamental domain, we obtain that $\mathrm{Tw}([\alpha^{-1}, \beta^{-1}]), p) = \pm 2\pi$, sign depending on orientation. This is as it should be, since the developing image of the boundary should be the axis of its holonomy; and we may conclude that $[\alpha, \beta] \in \mathrm{Hyp}_{\pm 1}$.

More generally, whenever we have a pentagon $\mathcal{P}(\alpha, \beta; p)$ which bounds an immersed disc, the pentagon extends to a developing map for a hyperbolic cone-manifold structure on S with piecewise geodesic boundary and one corner point; and every punctured torus with a hyperbolic cone-manifold structure with no interior cone points and at most one corner point can be cut into a pentagon in this way. For a complete investigation of such hyperbolic cone-manifold structures on punctured tori and their holonomy representations, see [14].

Consider again the twist of a commutator; it may be expanded as a product, using composition lemma 3.2:

$$\mathrm{Tw}\left([\tilde{\alpha}^{-1}, \tilde{\beta}^{-1}], p\right) = \mathrm{Tw}\left(\tilde{\beta}, p\right) + \mathrm{Tw}\left(\tilde{\alpha}, \beta p\right) + \mathrm{Tw}\left(\tilde{\beta}^{-1}, \alpha\beta p\right) + \mathrm{Tw}\left(\tilde{\alpha}^{-1}, \beta^{-1}\alpha\beta p\right) - \Delta[p_0, p_1, p_2, p_3, p_4],$$

where $\Delta[p_0, p_1, p_2, p_3, p_4]$ is the signed area of the pentagon formed by geodesic segments (giving oriented boundary) $p_0 \to p_1 \to p_2 \to p_3 \to p_4$. Clearly not both $[p_0, p_1, p_2, p_3, p_4]$ and $[p_0, p_4, p_1, p_2, p_3] = \mathcal{P}(\alpha, \beta; p)$ can bound embedded discs! If area is however understood as a Euclidean angle defect given by angles between succeeding segments, or by cutting into triangles and summing their signed area, the above is true in all cases. Setting $q = \alpha\beta p$ the four twists involved can be simplified:

$$\mathrm{Tw}\left(\tilde{\beta}, \beta^{-1}\alpha^{-1}q\right) + \mathrm{Tw}\left(\tilde{\alpha}, \alpha^{-1}q\right) + \mathrm{Tw}\left(\tilde{\beta}^{-1}, q\right) + \mathrm{Tw}\left(\tilde{\alpha}^{-1}, \beta^{-1}q\right)$$
$$= \mathrm{Tw}\left(\tilde{\beta}, \alpha^{-1}q\right) + \mathrm{Tw}\left(\tilde{\alpha}, q\right) - \mathrm{Tw}\left(\tilde{\beta}, q\right) - \mathrm{Tw}\left(\tilde{\alpha}, \beta^{-1}q\right).$$

Here we have used equation (3.1) and lemma 3.4. This can be considered as some kind of "twist cross ratio", the change in the twist of $\tilde{\alpha}$ under translation by β, relative to the change in twist of $\tilde{\beta}$ under translation by α. With proposition 3.7, these remarks give

$$\left(\mathrm{Tw}\left(\tilde{\alpha}, q\right) - \mathrm{Tw}\left(\tilde{\alpha}, \beta^{-1}q\right)\right) - \left(\mathrm{Tw}\left(\tilde{\beta}, q\right) - \mathrm{Tw}\left(\tilde{\beta}, \alpha^{-1}q\right)\right) = \Delta[p_0, p_1, p_2, p_3, p_4] + \Delta[p_0, p_4, p_1, p_2, p_3].$$

3.6   Commutators and twist bounds

Since a hyperbolic pentagon has area $< 3\pi$, proposition 3.7 implies that $\left|\mathrm{Tw}\left([\tilde{\alpha}, \tilde{\beta}], p\right)\right| < 3\pi$ when $\mathcal{P}(\alpha, \beta; p)$ is sufficiently nice. Such an inequality is true in general, as we now see.

Using the addition lemma 3.5 gives two expressions for $\mathrm{Tw}(\tilde{\alpha}, p) + \mathrm{Tw}\left(\tilde{\beta}, p\right)$:

$$\mathrm{Tw}(\tilde{\alpha}, p) + \mathrm{Tw}\left(\tilde{\beta}, p\right) = \mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) + \Delta[p, \beta p, \beta\alpha p] = \mathrm{Tw}\left(\tilde{\alpha}\tilde{\beta}, p\right) + \Delta[p, \alpha p, \alpha\beta p].$$

It follows that

$$\mathrm{Tw}\left(\tilde{\alpha}\tilde{\beta}, p\right) - \mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) = \Delta[p, \beta p, \beta\alpha p] - \Delta[p, \alpha p, \alpha\beta p].$$

Applying the addition and inverse lemmas to the commutator $[\tilde{\alpha}, \tilde{\beta}] = \left(\tilde{\alpha}\tilde{\beta}\right)\left(\tilde{\beta}\tilde{\alpha}\right)^{-1}$, we have

$$\mathrm{Tw}\left([\tilde{\alpha}, \tilde{\beta}], p\right) = \mathrm{Tw}\left(\tilde{\alpha}\tilde{\beta}, p\right) - \mathrm{Tw}\left(\tilde{\beta}\tilde{\alpha}, p\right) - \Delta\left[p, \alpha\beta p, [\alpha, \beta]p\right]$$

Putting these together, we immediately have the following.

Lemma 3.9 (Commutator area) *For any $\tilde{\alpha}, \tilde{\beta} \in \widetilde{PSL_2\mathbb{R}}$ and $p \in \overline{\mathbb{H}}^2$,*

$$\mathrm{Tw}\left(\left[\tilde{\alpha}, \tilde{\beta}\right], p\right) = \Delta[p, \beta p, \beta\alpha p] - \Delta[p, \alpha p, \alpha\beta p] - \Delta\left[p, \alpha\beta p, [\alpha, \beta]p\right] \qquad\blacksquare$$

Hence for any $\tilde{\alpha}, \tilde{\beta} \in \widetilde{PSL_2\mathbb{R}}$ and $p \in \mathbb{H}^2$, we have (for $p \in S^1_\infty$ the inequality is not strict):

$$\left|\mathrm{Tw}\left(\left[\tilde{\alpha}, \tilde{\beta}\right], p\right)\right| < 3\pi. \tag{3.3}$$

We can say more about the possible values for commutators; we consider the elliptic, parabolic, identity and hyperbolic cases separately.

If $[\alpha, \beta]$ is elliptic then take $p = \mathrm{Fix}[\alpha, \beta]$. The triangle formed by $p, \alpha\beta p, [\alpha, \beta]p$ then has zero area, and the twist of $[\tilde{\alpha}, \tilde{\beta}]$ at this point is $< 2\pi$ in magnitude. Thus $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Ell}_{-1} \cup \mathrm{Ell}_1$.

If $[\alpha, \beta]$ is parabolic, set $p = \mathrm{Fix}[\alpha, \beta] \in S^1_\infty$; then again $\Delta\left[p, \alpha\beta p, [\alpha\beta]p\right] = 0$, and so $|\mathrm{Tw}\left([\alpha, \beta], p\right)| \le 2\pi$. In fact this twist must be in $\{-2\pi, 0, 2\pi\}$, and $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Par}_{-1} \cup \mathrm{Par}_0 \cup \mathrm{Par}_1$ respectively.

We can say something more in this case, with a little more work. Suppose that $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Par}_1$; we will show in fact it lies in Par_1^-. We have $\mathrm{Tw}\left([\alpha, \beta], p\right) = \Delta[p, \beta p, \beta\alpha p] - \Delta[p, \alpha p, \alpha\beta p]$. Applying the isometries $\alpha^{-1}\beta^{-1}$ and $\beta^{-1}\alpha^{-1}$ respectively to these ideal triangles gives this twist as $\Delta[\alpha^{-1}\beta^{-1}p, \alpha^{-1}p, p] - \Delta[\beta^{-1}\alpha^{-1}p, \beta^{-1}p, p]$; these are both ideal triangles, and the values are $\pm\pi$ according to orientation, or 0 if degenerate. Note that since $p = \mathrm{Fix}[\alpha, \beta]$, $\alpha^{-1}\beta^{-1}p = \beta^{-1}\alpha^{-1}p$. The only way to obtain 2π for the twist (and hence to lie in Par_1), then, is if the four points $p, \beta^{-1}p, \alpha^{-1}\beta^{-1}p = \beta^{-1}\alpha^{-1}p, \alpha^{-1}p$ lie in anticlockwise order around the circle. Since α takes $(\alpha^{-1}p, \alpha^{-1}\beta^{-1}p) \mapsto (p, \beta^{-1}p)$, in different directions around S^1_∞, α must be hyperbolic; similarly for β, and their axes must cross, and lie as shown in figure 7. Let x be the endpoint of $\mathrm{Axis}\,\beta$ shown; we now chase x around the diagram to $[\alpha, \beta]x$. First $\beta^{-1}x = x$. As $\beta^{-1}x$ is anticlockwise of $\beta^{-1}p$, then $\alpha\beta^{-1}x$ must lie on the same side of $\mathrm{Axis}\,\alpha$ and anticlockwise of $\alpha^{-1}\beta^{-1}p$, hence as shown. Then $\alpha^{-1}\beta^{-1}x$ is anticlockwise of $\beta^{-1}\alpha^{-1}p$, so $\beta\alpha^{-1}\beta^{-1}x$ must lie anticlockwise of $\alpha^{-1}p$ and on the same side of $\mathrm{Axis}\,\beta$, hence as shown. By similar reasoning $[\alpha, \beta]x$ lies on the same side of $\mathrm{Axis}\,\alpha$ as $\beta\alpha^{-1}\beta^{-1}x$ and anticlockwise of p. Considering the result of applying $[\alpha, \beta]$ to x, we conclude that $[\alpha, \beta] \in \mathrm{Par}_1^-$.

If $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Par}_{-1}$, then the same argument applies and in fact it lies in Par_{-1}^+.

If $[\alpha, \beta]$ is the identity, then we see that $[\tilde{\alpha}, \tilde{\beta}]$ is the identity in $\widetilde{PSL_2\mathbb{R}}$. If either of α or β is the identity this is immediate; if not, α or β are of the same type with same fixed points; and hence taking lifts and following unit vectors it is clear.
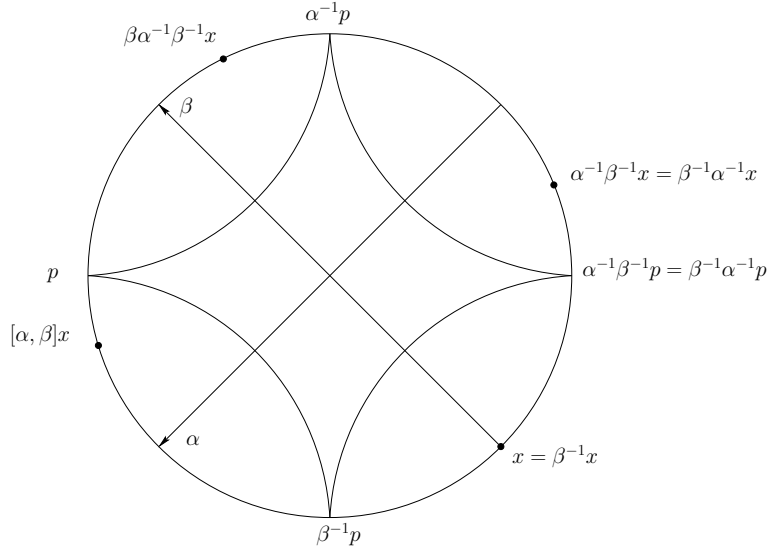
If $[\alpha, \beta]$ is hyperbolic, then equation (3.3) immediately gives $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Hyp}_{-1} \cup \mathrm{Hyp}_0 \cup \mathrm{Hyp}_1$. As it turns out, these are all possible.

We have now proved the following theorem, which appears in [20, 7, 10]; we also give a different proof in [13]. See also figure 8.

Theorem 3.10 *For $\tilde{\alpha}, \tilde{\beta} \in \widetilde{PSL_2\mathbb{R}}$,*

$$\left[\tilde{\alpha}, \tilde{\beta}\right] \in \{1\} \cup \left(\bigcup_{n=-1}^{1} \mathrm{Hyp}_n \cup \mathrm{Ell}_n\right) \cup \mathrm{Par}_0 \cup \mathrm{Par}_{-1}^+ \cup \mathrm{Par}_1^-. \qquad\blacksquare$$

Here we take $\mathrm{Ell}_0 = \emptyset$ for convenience. Combining this proposition with the trace lemma 2.2 gives an immediate corollary.

Figure 7: Arrangement of axes in the case $[\alpha, \beta] \in \mathrm{Par}_1$.

Corollary 3.11 *If $\alpha, \beta \in PSL_2\mathbb{R}$ then*

(i) $\mathrm{Tr}[\alpha, \beta] > 2$ *implies* $[\alpha, \beta] \in \mathrm{Hyp}_0$;

(ii) $\mathrm{Tr}[\alpha, \beta] = 2$ *implies* $[\alpha, \beta] \in \{1\} \cup \mathrm{Par}_0$;

(iii) $\mathrm{Tr}[\alpha, \beta] \in (-2, 2)$ *implies* $[\alpha, \beta] \in \mathrm{Ell}_{-1} \cup \mathrm{Ell}_1$;

(iv) $\mathrm{Tr}[\alpha, \beta] = -2$ *implies* $[\alpha, \beta] \in \mathrm{Par}^+_{-1} \cup \mathrm{Par}^-_1$;

(v) $\mathrm{Tr}[\alpha, \beta] < -2$ *implies* $[\alpha, \beta] \in \mathrm{Hyp}_{-1} \cup \mathrm{Hyp}_1$.

∎

3.7 Commutators and arrangements of axes

In the proof of proposition 3.10 we showed that if $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Par}^{\mp}_{\pm 1}$ then α, β are hyperbolic and their axes cross. We continue such analysis for other commutators, and make conclusions about the type and location of $\tilde{\alpha}, \tilde{\beta}$. In particular we prove the following result, which appears in [11].
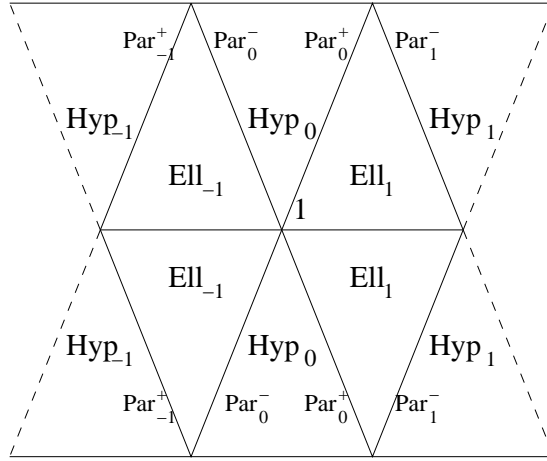
Proposition 3.12 (Goldman [11]) *Let $\alpha, \beta \in PSL_2\mathbb{R}$. The following are equivalent:*

(i) α, β *are hyperbolic and their axes cross;*

(ii) $[\alpha, \beta] \in \mathrm{Ell}_{\pm 1} \cup \mathrm{Par}^{\mp}_{\pm 1} \cup \mathrm{Hyp}_{\pm 1}$;

(iii) $\mathrm{Tr}[\alpha, \beta] < 2$.

We use a couple of lemmas. The first was implicitly used in the argument of the previous section and is straightforward. The second is a simple computation, for instance using Fermi coordinates (see e.g. [5] p. 38).

Lemma 3.13 *Let A, B, C, D be points on S^1_∞ in anticlockwise order. Suppose $\alpha \in PSL_2\mathbb{R}$ is an isometry which takes A to B and D to C. Then α is hyperbolic; the repulsive fixed point of α lies in the interval of S^1_∞ between D and A; and the attractive fixed point between B and C.* ∎

Figure 8: Possible commutators in $\widetilde{PSL_2}\mathbb{R}$.

Lemma 3.14 *Let $\alpha \in PSL_2\mathbb{R}$ be hyperbolic and $x_1, y_1 \in \mathbb{H}^2$. The translation distances of α at x_1 and x_2 are equal, i.e. $d(x_1, \alpha x_1) = d(x_2, \alpha x_2)$, if and only if x_1, x_2 lie at the same perpendicular distance from $\mathrm{Axis}\,\alpha$.* ∎

PROOF (OF PROPOSITION 3.12) The equivalence of (ii) and (iii) is immediate from corollary 3.11. To prove (ii) implies (i), we consider the various possible cases for $[\tilde{\alpha}, \tilde{\beta}]$.

- $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Hyp}_{\pm 1}$. The argument is virtually identical to the $\mathrm{Par}_{\pm 1}^{\mp}$ case. Consider $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Hyp}_1$; the case Hyp_{-1} is identical with reversed orientation. Apply lemma 3.9 taking $p \in S_\infty^1$ to be a fixed point of $[\tilde{\alpha}, \tilde{\beta}]$; so $\alpha^{-1}\beta^{-1}p = \beta^{-1}\alpha^{-1}p$ and, since $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Hyp}_1$, $\mathrm{Tw}([\tilde{\alpha}, \tilde{\beta}], p) = 2\pi$. Note that $\Delta[p, \beta p, \beta \alpha p] = \Delta[\alpha^{-1}\beta^{-1}p, \alpha^{-1}p, p]$ and $\Delta[p, \alpha p, \alpha \beta p] = \Delta[\beta^{-1}\alpha^{-1}p, \beta^{-1}p, p]$. Then $p, \beta^{-1}p, \alpha^{-1}\beta^{-1}p = \beta^{-1}\alpha^{-1}p, \alpha^{-1}p$ must occur in anticlockwise order around S_∞^1. By two applications of lemma 3.13 then α, β are hyperbolic and their axes cross.

- $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Par}_{\pm 1}^{\mp}$. We considered this case above, and concluded α, β hyperbolic with axes crossing.

- $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Ell}_{\pm 1}$. Consider $[\tilde{\alpha}, \tilde{\beta}] \in \mathrm{Ell}_1$; the Ell_{-1} case is identical with reversed orientation. Apply lemma 3.9 as above, taking $p = \mathrm{Fix}[\alpha, \beta]$. So $\alpha^{-1}\beta^{-1}p = \beta^{-1}\alpha^{-1}p$ and we have $\mathrm{Tw}([\tilde{\alpha}, \tilde{\beta}], p) = \Delta[\alpha^{-1}\beta^{-1}p, \alpha^{-1}p, p] - \Delta[\beta^{-1}\alpha^{-1}p, \beta^{-1}p, p]$. These two triangles are *congruent*: they both contain the side $p \to \alpha^{-1}\beta^{-1}p$, the isometry α takes the side $\alpha^{-1}\beta^{-1}p \to \alpha^{-1}p$ to $\beta^{-1}p \to p$, and the isometry β takes the side $\beta^{-1}\alpha^{-1}p \to \beta^{-1}p$ to $\alpha^{-1}p \to p$. As $\mathrm{Tw}([\tilde{\alpha}, \tilde{\beta}], p) > 0$, the two congruent triangles fit together to give a non-self-intersecting quadrilateral Q formed with vertices, in anticlockwise order, $p, \beta^{-1}p, \alpha^{-1}\beta^{-1}p = \beta^{-1}\alpha^{-1}p, \alpha^{-1}p$. Being constructed out of two congruent triangles, the opposite interior angles of Q are equal. Moreover, extending the four side segments out to infinity, the only intersection points are the four vertices of the quadrilateral.

  Extending the opposite sides $p \to \beta^{-1}p$ and $\alpha^{-1}p \to \alpha^{-1}\beta^{-1}p$ to infinity, then, they do not intersect; and they are related by α. By lemma 3.13 then α is hyperbolic, and $\mathrm{Axis}\,\alpha$ has fixed points at infinity separated by these lines. In particular, $\mathrm{Axis}\,\alpha$ intersects $p \to \beta^{-1}p$ (possibly extended) at a single point, and intersects $\alpha^{-1}p \to \alpha^{-1}\beta^{-1}p$ (possibly extended) at a single point also. Moreover, since $\alpha^{-1}p \to p$ and $\beta^{-1}\alpha^{-1}p \to \beta^{-1}p$ have the same length (being related by β), the distance between x and $\alpha(x)$ is the same for $x = \alpha^{-1}p$ and $x = \alpha^{-1}\beta^{-1}p$. By lemma 3.14, $\alpha^{-1}p$ and $\alpha^{-1}\beta^{-1}p$ lie at the same perpendicular distance from $\mathrm{Axis}\,\alpha$. Hence they lie on opposite sides of $\mathrm{Axis}\,\alpha$, and $\mathrm{Axis}\,\alpha$ intersects the segment $\alpha^{-1}p \to \beta^{-1}\alpha^{-1}p$. By the same argument regarding translation distances of α^{-1}, $\mathrm{Axis}\,\alpha^{-1} = \mathrm{Axis}\,\alpha$ intersects the segment $p \to \beta^{-1}p$. That is, $\mathrm{Axis}\,\alpha$ intersects two opposite sides of the quadrilateral Q. By the same

14

argument, β is hyperbolic and $\operatorname{Axis}\beta$ intersects the other pair of opposite sides of Q. Hence $\operatorname{Axis}\alpha$ and $\operatorname{Axis}\beta$ intersect.

To prove (i) implies (iii), we repeat the argument of [11], writing matrices for $\alpha, \beta \in PSL_2\mathbb{R}$. We may conjugate so that α has fixed points $-1, 1$ in the upper half plane, and β has fixed points r, ∞ where $r \in (-1, 1)$. We may write

$$\alpha = \begin{pmatrix} \cosh x & \sinh x \\ \sinh x & \cosh x \end{pmatrix}, \quad \beta = \begin{pmatrix} e^y & -2r\sinh y \\ 0 & e^{-y} \end{pmatrix}$$

where $x, y \in \mathbb{R}$. A calculation then gives

$$\operatorname{Tr}[\alpha, \beta] = 2 + 4\left(r^2 - 1\right)\sinh^2 x \sinh^2 y < 2. \qquad \blacksquare$$

Note Goldman's proof in [11] of this proposition is entirely algebraic; the difference here is that we have proved one direction geometrically.

3.8   Surface group representations and the Milnor–Wood inequality

Now consider $\tilde{\alpha}_1, \tilde{\beta}_1, \ldots, \tilde{\alpha}_g, \tilde{\beta}_g, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_n \in \widetilde{PSL_2\mathbb{R}}$ and consider

$$\left[\tilde{\alpha}_1, \tilde{\beta}_1\right] \left[\tilde{\alpha}_2, \tilde{\beta}_2\right] \cdots \left[\tilde{\alpha}_g, \tilde{\beta}_g\right] \tilde{\gamma}_1 \tilde{\gamma}_2 \cdots \tilde{\gamma}_n.$$

The commutators, as we saw in lemma 2.1, are independent of choice of lift of α_i, β_i; for the $\tilde{\gamma}_n$, let us assume they are "efficiently chosen", i.e. have twist less than π in magnitude at some point p. Such an expression is of course the relator in the standard presentation of the fundamental group of the surface S of genus g with n boundary components. Such $\alpha_i, \beta_i, \gamma_i \in PSL_2\mathbb{R}$ arise as the holonomy of a hyperbolic structure on S, and in this case the product is $1 \in PSL_2\mathbb{R}$; it therefore lifts to some $\mathbf{z}^m \in \widetilde{PSL_2\mathbb{R}}$. We ask how large m can be, i.e. how large the twist of the relator expression can be.

Repeated use of the addition lemma gives a bound, in the "hyperbolic" case $\chi(S) = 2 - 2g - n < 0$.

Theorem 3.15 (Milnor [18]) *Let $p \in \mathbb{H}^2$; let $\tilde{\alpha}_1, \tilde{\beta}_1, \ldots, \tilde{\alpha}_g, \tilde{\beta}_g, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_n \in \widetilde{PSL_2\mathbb{R}}$ with $2 - 2g - n < 0$ and $|\operatorname{Tw}(\tilde{\gamma}_i, p)| \leq \pi$; assume*

$$[\alpha_1, \beta_1]\,[\alpha_2, \beta_2]\,\cdots\,[\alpha_g, \beta_g]\,\gamma_1\gamma_2\cdots\gamma_n = 1 \in PSL_2\mathbb{R}.$$

Then

$$\left[\tilde{\alpha}_1, \tilde{\beta}_1\right]\,\left[\tilde{\alpha}_2, \tilde{\beta}_2\right]\,\cdots\,\left[\tilde{\alpha}_g, \tilde{\beta}_g\right]\,\tilde{\gamma}_1\tilde{\gamma}_2\cdots\tilde{\gamma}_n = \mathbf{z}^m$$

where $|m| \leq |\chi(S)| = 2g + n - 2$.

PROOF There are $4g + n$ terms in the expression; we use the addition lemma $4g + n - 1$ times. Note that whenever $\tilde{\xi}\tilde{\eta} = 1$ we have $\Delta[p, \xi p, \xi\eta p] = 0$, being a degenerate triangle. So the first time we use the addition lemma the triangle is degenerate, and there are at most $4g + n - 2$ nondegenerate triangles. Once we have used the addition lemma $4g + n - 1$ times, we have the difference between the twist of the relator and the twist of the individual terms as a sum of signed areas of $4g + n - 2$ triangles; denote these signed areas $\Delta_1, \ldots, \Delta_{4g+n-2}$. The inverses from the commutators cancel, thanks to lemma 3.4, leaving only the twists of the $\tilde{\gamma}_i$. Thus

$$\operatorname{Tw}(\mathbf{z}^m, p) = \sum_{i=1}^{n} \operatorname{Tw}(\tilde{\gamma}_i, p) + \sum_{i=1}^{4g+n-2} \Delta_i,$$

and since triangle areas are $< \pi$ and the twists of the $\tilde{\gamma}_i$ are by assumption $\leq \pi$,

$$2\pi\,|m| = \left|\operatorname{Tw}(\mathbf{z}^m, p)\right| < n\pi + (4g + n - 2)\pi = 2\pi\,(2g + n - 1).$$

Since $|m| < 2g + n - 1$ and m is an integer, we are done. $\qquad \blacksquare$

Note that the above argument does not work for the "non-hyperbolic" case $2g+n \leq 2$. For instance, with $k=0, n=2$, setting $\tilde{c}_1 = \tilde{c}_2$ to be half turns of twist π, we have $c_1 c_2 = 1 \in PSL_2\mathbb{R}$ but $\tilde{c}_1\tilde{c}_2 = \mathbf{z}$.

Suppose we have a representation $\rho : \pi_1(S) \longrightarrow PSL_2\mathbb{R}$. After choosing lifts $\tilde{\gamma}_i$ of the images of the boundary components, we have a lift of the image of the relator, which is some $\mathbf{z}^m$. This m is essentially the (relative) *Euler class* $\mathcal{E}(\rho)$ of the representation: see e.g. [10, 15]. More precisely $\mathcal{E}(\rho) \in H^2(S)$ and takes the fundamental class $[S]$ to m. Interpreting the above result this way we have:

Theorem 3.16 *Let $\rho : \pi_1(S) \longrightarrow PSL_2\mathbb{R}$ be a representation. Taking lifts $\tilde{\gamma}_i$ of the images of the boundary components with $|\mathrm{Tw}(\tilde{\gamma}_i, p)| < \pi$ at some point p, the (relative) Euler class $\mathcal{E}(\rho)$ takes the fundamental class $[S]$ to $m \in \mathbb{Z}$, where $|m| \leq |\chi(S)|$.* ∎

The above two results, and similar formulations, are known generally as the *Milnor-Wood inequality*. The first inequality was first proved by Milnor [18], generalised by Wood [20], reproved by Goldman [10], and generalised further by Eisenbud–Hirsch–Neumann [7]. The reformulation in terms of Euler class was first given, so far as we know, by Wood [20].

3.9 Larger products and polygons

Using our existing results — composition lemma 3.2, addition lemma 3.5, commutator pentagon 3.7, and commutator area 3.9 — together, we obtain results for more complicated expressions and figures.

For instance, using the composition lemma repeatedly on a product $\tilde{\gamma}_1, \tilde{\gamma}_2 \cdots \tilde{\gamma}_n$ gives

$$\mathrm{Tw}\left(\prod_{i=1}^{n}\tilde{\gamma}_i, p\right) = \sum_{i=1}^{n}\mathrm{Tw}\left(\tilde{\gamma}_i, \left(\prod_{j=i+1}^{n}\gamma_j\right)p\right) - \sum_{i=2}^{n}\Delta\left[\left(\prod_{j=i+1}^{n}\gamma_j\right)p, \left(\prod_{j=i}^{n}\gamma_j\right)p, \left(\prod_{j=1}^{n}\gamma_j\right)p\right].$$

But these triangles share successive sides. Consider the polygon $\mathcal{C}(\gamma; p)$ with $n+1$ sides

$$p \to \gamma_1\gamma_2\cdots\gamma_n p \to \gamma_2\cdots\gamma_n p \to \cdots \to \gamma_{n-1}\gamma_n p \to \gamma_n p \to p,$$

oriented by the direction on the boundary above. If $\mathcal{C}(\gamma; p)$ is convex, then it can be cut into precisely the triangles occurring in the above sum; and so, as long as the polygon is simple (i.e. non-self-intersecting), the above expression gives the area.

Lemma 3.17 (Composition polygon) *Let $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_n \in \widetilde{PSL_2\mathbb{R}}$ and $p \in \overline{\mathbb{H}}^2$. Suppose the polygon $\mathcal{C}(\gamma_1, \ldots, \gamma_n; p)$ is simple. Then*

$$\mathrm{Tw}\left(\prod_{i=1}^{n}\tilde{\gamma}_i, p\right) = \sum_{i=1}^{n}\mathrm{Tw}\left(\tilde{\gamma}_i, \left(\prod_{j=i+1}^{n}\gamma_j\right)p\right) + \Delta\left[\mathcal{C}\left(\gamma_1, \ldots, \gamma_n; p\right)\right].$$ ∎

Note that if $\gamma_1\cdots\gamma_n = 1 \in PSL_2\mathbb{R}$ then $\mathcal{C}(\gamma_1, \ldots, \gamma_n; p)$ reduces to an n-gon and the result still holds.

Consider now polygons of the type arising as fundamental domains of hyperbolic surfaces. Given $\tilde{\alpha}_1, \tilde{\beta}_1, \cdots, \tilde{\alpha}_g, \tilde{\beta}_g, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_n$, we consider a polygon $\mathcal{D}(\alpha, \beta, \gamma; p)$ associated to the surface group relator

$$\left[\tilde{\alpha}_1, \tilde{\beta}_1\right]\left[\tilde{\alpha}_2, \tilde{\beta}_2\right]\cdots\left[\tilde{\alpha}_g, \tilde{\beta}_g\right]\tilde{\gamma}_1\tilde{\gamma}_2\cdots\tilde{\gamma}_n.$$

The vertices of $\mathcal{D}(\alpha, \beta, \gamma; p)$ are, along the oriented boundary:

$$p \to [\alpha_1, \beta_1] \cdots [\alpha_g, \beta_g] \gamma_1 \cdots \gamma_n p$$
$$\to \beta_1^{-1}[\alpha_2, \beta_2] \cdots [\alpha_g, \beta_g] \gamma_1 \cdots \gamma_n p \to \alpha_1^{-1}\beta_1^{-1}[\alpha_2, \beta_2] \cdots \gamma_n p \to \beta_1 \alpha_1^{-1} \beta_1^{-1}[\alpha_2, \beta_2] \cdots \gamma_n p$$
$$\to [\alpha_2, \beta_2] \cdots [\alpha_g, \beta_g] \gamma_1 \cdots \gamma_n p$$
$$\to \beta_2^{-1}[\alpha_3, \beta_3] \cdots [\alpha_g, \beta_g] \gamma_1 \cdots \gamma_n p \to \alpha_2^{-1}\beta_2^{-1}[\alpha_3, \beta_3] \cdots \gamma_n p \to \beta_2 \alpha_2^{-1} \beta_2^{-1}[\alpha_3, \beta_3] \cdots \gamma_n p$$
$$\to [\alpha_3 \beta_3] \cdots [\alpha_g, \beta_g] \gamma_1 \cdots \gamma_n p$$
$$\to \cdots$$
$$\to \gamma_1 \cdots \gamma_n p$$
$$\to \gamma_2 \cdots \gamma_n p \to \gamma_3 \cdots \gamma_n p \to \cdots \to \gamma_n p$$
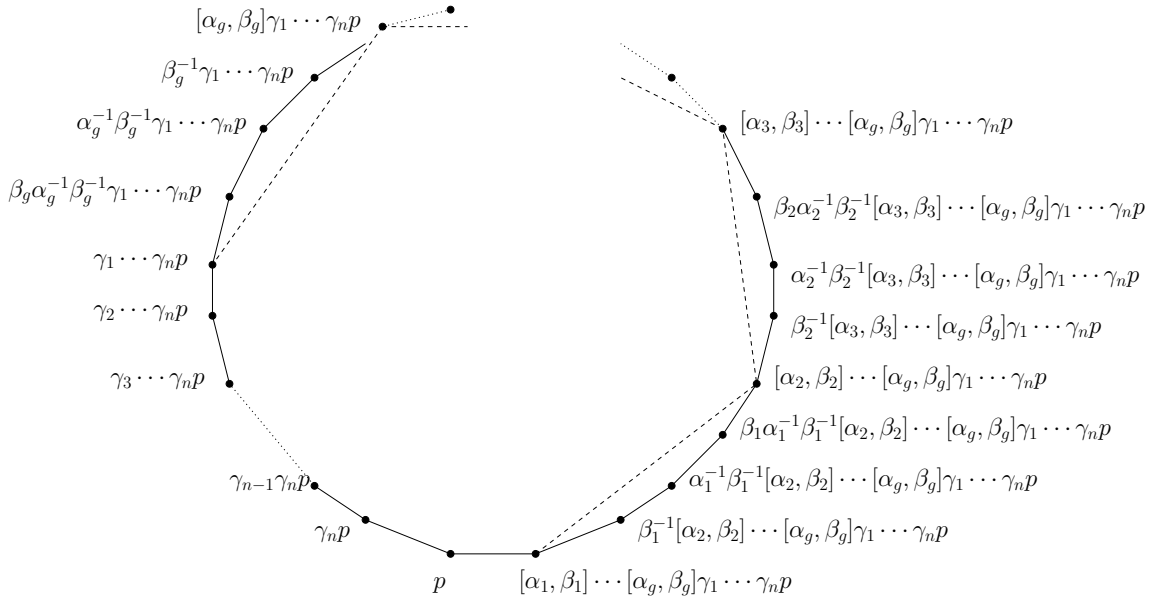$$\to p$$

See figure 9.

Figure 9: The polygon $\mathcal{D}(\alpha, \beta, \gamma; p)$.

Note that $\mathcal{D}(\alpha, \beta, \gamma; p)$ can be considered as a sequence of pentagons attached to the polygon

$$\mathcal{C}([\alpha_1, \beta_1], [\alpha_2, \beta_2], \ldots, [\alpha_g, \beta_g], \gamma_1, \ldots, \gamma_n; p).$$

(See the dashed lines in figure 9.) Note that $\mathcal{D}(\alpha, \beta, \gamma; p)$ is a $(4g + n + 1)$-gon, but in a surface group representation $[\alpha_1, \beta_1] \cdots [\alpha_g, \beta_g] \gamma_1 \cdots \gamma_n = 1$, so it is a $(4g + n)$-gon. In any case, applying the composition polygon lemma 3.17 gives

$$\mathrm{Tw}\left(\prod_{i=1}^{g} \left[\tilde{\alpha}_i, \tilde{\beta}_i \right] \prod_{i=1}^{n} \tilde{\gamma}_i, p \right) = \sum_{i=1}^{g} \mathrm{Tw}\left(\left[\tilde{\alpha}_i, \tilde{\beta}_i \right], \left(\prod_{j=i+1}^{g} [\alpha_j, \beta_j] \right) \left(\prod_{j=1}^{n} \gamma_j \right) p \right)$$
$$+ \sum_{i=1}^{n} \mathrm{Tw}\left(\tilde{\gamma}_i, \left(\prod_{j=i+1}^{n} \tilde{\gamma}_j \right) p \right) + \Delta \left[\mathcal{C}\left([\alpha_1, \beta_1], \ldots, [\alpha_g, \beta_g], \gamma_1, \ldots, \gamma_n; p \right) \right].$$

But the twist of a commutator is the area of a pentagon (proposition 3.7), hence the above is equal to

$$\sum_{i=1}^{n} \mathrm{Tw} \left(\tilde{\gamma}_i, \left(\prod_{j=i+1}^{n} \tilde{\gamma}_j \right) p \right) + \Delta \left[\mathcal{C} \left([\alpha_1, \beta_1], \ldots, [\alpha_g, \beta_g], \gamma_1, \ldots, \gamma_n; p \right) \right]$$

$$+ \sum_{i=1}^{g} \Delta \left[\mathcal{P} \left(\alpha_i^{-1}, \beta_i^{-1}; \left(\prod_{j=i+1}^{g} [\alpha_j, \beta_j] \right) \left(\prod_{j=1}^{n} \gamma_j \right) p \right) \right]$$

These areas simply add up to our polygon $\mathcal{D}(\alpha, \beta, \gamma; p)$ — provided that we have simple polygons. In order to ensure that the entire polygon, and the result of cutting off these pentagons, are simple, the easiest thing to do is make an assumption of *convexity*.

Theorem 3.18 *Let* $\tilde{\alpha}_1, \tilde{\beta}_1, \cdots, \tilde{\alpha}_g, \tilde{\beta}_g, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_n \in \widetilde{PSL_2\mathbb{R}}$ *and* $p \in \overline{\mathbb{H}}^2$. *Suppose that* $\mathcal{D}(\alpha, \beta, \gamma; p)$ *is a convex polygon. Then*

$$\mathrm{Tw} \left(\left[\tilde{\alpha}_1, \tilde{\beta}_1 \right] \cdots \left[\tilde{\alpha}_g, \tilde{\beta}_g \right] \tilde{\gamma}_1 \cdots \tilde{\gamma}_n, p \right) = \sum_{i=1}^{n} \mathrm{Tw} \left(\tilde{\gamma}_i, \gamma_{i+1} \gamma_{i+2} \cdots \gamma_n p \right) + \Delta \left[\mathcal{D} \left(\alpha, \beta, \gamma; p \right) \right]. \qquad \blacksquare$$

This gives another proof of the Milnor–Wood inequality, provided $\mathcal{D}(\alpha, \beta, \gamma; p)$ is convex: if $\rho : \pi_1(S) \longrightarrow PSL_2\mathbb{R}$ is a representation and $\mathcal{E}(\rho) = m$ then

$$2\pi m = \sum_{i=1}^{n} \mathrm{Tw} \left(\tilde{\gamma}_i, \gamma_{i+1} \gamma_{i+2} \cdots \gamma_n p \right) + \Delta \left[\mathcal{D} \left(\alpha, \beta, \gamma; p \right) \right]$$

As before, we choose lifts $\tilde{\gamma}_i$ of the boundary components with twists less than π in magnitude. Noting that the area of a hyperbolic $(4g + n)$-gon is less than $(4g + n - 2)\pi$, we immediately have

$$2\pi |m| < n\pi + \pi (4g + n - 2) = 2\pi (2g + n - 1),$$

and again as m is an integer, $|m| \leq 2g + n - 2 = |\chi(S)|$.

4 Milnor's angle function

4.1 Definition

We now recall the definition of Milnor's angle function, relate it to our notion of twist, and then deduce various properties from our results on twisting.

A matrix $\alpha \in GL_2^+\mathbb{R}$ can be written uniquely in the form $\alpha = R(\alpha)S(\alpha)$, where $R(\alpha), S(\alpha) \in GL_2^+\mathbb{R}$, R is orthogonal (i.e. $R \in SO_2\mathbb{R}$) and S is symmetric positive definite. Since R is orthogonal, it is of the form

$$R(\alpha) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

for some θ. This θ can be thought of as the angle of rotation of α; and hence we have a function $\theta : GL_2^+\mathbb{R} \longrightarrow \mathbb{R}/2\pi\mathbb{Z}$. Milnor's Θ function is a lift of this map, $\Theta : \widetilde{GL_2^+\mathbb{R}} \longrightarrow \mathbb{R}$.

Indeed, the map $R : SL_2\mathbb{R} \longrightarrow SO_2\mathbb{R} \cong S^1$ is a retraction, which lifts to a retraction $\tilde{R} : \widetilde{SL_2\mathbb{R}} \longrightarrow \widetilde{SO_2\mathbb{R}} \cong \mathbb{R}$. Since $SO_2\mathbb{R} \cong S^1$, we have $\widetilde{SO_2\mathbb{R}} \cong \mathbb{R}$, $\mathfrak{so}_2\mathbb{R} \cong \mathbb{R}$, so the exponential map is

$$\exp : \mathbb{R} \longrightarrow SO_2\mathbb{R}, \qquad \theta \mapsto \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

lifts to $\widetilde{\exp} : \mathbb{R} \longrightarrow \widetilde{SO_2\mathbb{R}} \subset \widetilde{PSL_2\mathbb{R}}$. We define the angle function $\Theta : \widetilde{PSL_2\mathbb{R}} \longrightarrow \mathbb{R}$ by

$$\Theta(\tilde{\alpha}) = \widetilde{\exp}^{-1}\left(\tilde{R}(\alpha)\right).$$

Although Milnor considers Θ on $\widetilde{GL_2^+\mathbb{R}}$ and general 2×2 real matrices with positive determinant, we only consider $\widetilde{SL_2\mathbb{R}}$ and 2×2 real matrices with determinant 1. But any matrix with positive determinant can act as a hyperbolic isometry on the upper half plane as a fractional linear transformation; indeed, as groups, $GL_2^+\mathbb{R} \cong \mathbb{R}_+ \times SL_2\mathbb{R}$, under the isomorphism $A \mapsto (\det A, A/\det A)$, and the fractional linear transformations of A and λA are equal for any $\lambda \in \mathbb{R}_+$. We see that $\Theta(\lambda A) = \Theta(A)$. So restricting to $SL_2\mathbb{R}$ in essence loses no generality; if we like we could define twist on $\widetilde{GL_2^+\mathbb{R}} \times \overline{\mathbb{H}}^2$ without any difficulty.

4.2   A geometric interpretation

We now interpret Milnor's Θ through twisting. Although it seems that similar ideas have been used previously, for instance in [10, 20], as far as we know this has not been described explicitly before.

Proposition 4.1 *Let $\tilde{\alpha} \in \widetilde{PSL_2\mathbb{R}}$. Then*

$$\Theta(\tilde{\alpha}) = \frac{1}{2}\operatorname{Tw}(\tilde{\alpha}, i).$$

For the proof, we begin with some simple observations.

Lemma 4.2 *In the upper half plane model, the geodesic with endpoints at infinity $a, b \in \mathbb{R}$ passes through i if and only if $ab = -1$.* ∎

Lemma 4.3 *In the upper half plane model, the matrix*

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in SO_2\mathbb{R} \subset SL_2\mathbb{R},$$

taken modulo sign (i.e. in $PSL_2\mathbb{R} = \operatorname{Isom}^+ \mathbb{H}^2$), acts as a rotation of angle 2θ anticlockwise about i. ∎

Lemma 4.4 *An isometry of $\mathbb{H}^2$ is represented by a symmetric positive definite matrix in $SL_2\mathbb{R}$ other than the identity if and only if it is hyperbolic and its axis passes through i.*

PROOF A simple computation, contained in [13]. ∎

PROOF (OF PROPOSITION 4.1) From the definition of Θ we have

$$\tilde{\alpha} = \widetilde{\exp}\left(\Theta(\tilde{\alpha})\right)\tilde{S}(\tilde{\alpha})$$

where $\widetilde{\exp}(\Theta(\tilde{\alpha})) = \tilde{R}(\tilde{\alpha})$ and $\tilde{S}(\tilde{\alpha})$ is the lift of a symmetric positive definite matrix; since $\tilde{R}$ is a retraction, $\tilde{S}(\tilde{\alpha})$ is connected to the identity through lifts of symmetric positive definite matrices (hence lifts of hyperbolic isometries). Thus $\tilde{S}(\tilde{\alpha}) \in \operatorname{Hyp}_0 \cup \{1\}$.

Consider now the action of $\tilde{\alpha}$ on the hyperbolic plane. Since $S(\tilde{\alpha})$ is symmetric positive definite, by 4.4 it is either the identity, or a translation along an axis passing through i. And $\tilde{S}(\tilde{\alpha}) \in \operatorname{Hyp}_0 \cup \{1\}$ is the simplest lift of this translation. This action is followed by that of $\widetilde{\exp}(\Theta(\tilde{\alpha}))$, which is a rotation of angle $2\Theta(\tilde{\alpha})$ about i (by 4.3). So the overall action of $\tilde{\alpha}$ is to translate from $\alpha^{-1}(i)$ to i, and then rotate by angle $2\Theta(\tilde{\alpha})$. Thus $\operatorname{Tw}(\tilde{\alpha}, \alpha^{-1}(i)) = 2\Theta(\tilde{\alpha})$. By equation 3.1, this is also equal to $\operatorname{Tw}(\tilde{\alpha}, i)$. ∎

We note a comparison with Wood's methods in [20], which presage subsequent work on the Milnor–Wood inequality for homeomorphisms of the circle (e.g. [7, 6]). As mentioned above in section 2.4, Wood regards elements of $SL_2\mathbb{R}$ as acting on the S^1 of oriented lines through the origin in $\mathbb{R}^2$, which is equivalent to considering the action on the circle at infinity; we consider the action on the S^1 of unit tangent vectors at a point. Wood shows that his function $\widetilde{\mathrm{Top}^+ S^1} \longrightarrow \mathbb{R}$ is an extension of Milnor's Θ. We have shown that our twist at the particular point i is equal to Θ. Hence the action we consider on unit tangent vectors at i is equivalent to Wood's action on the circle at infinity. Although our function Tw is only defined on group $\widetilde{PSL_2}\mathbb{R}$, which is much smaller than $\widetilde{\mathrm{Top}^+ S^1}$, in applying to different points in $\mathbb{H}^2$ it is in a sense a generalisation of Wood's function.

4.3   Properties

We can now immediately deduce properties of Θ from Tw. Note several of our results for Tw, particularly involving surface group relators, cannot be expressed simply in terms of Θ, since they involve twists at different points. In this sense, the twist is a true generalisation of Θ.

Immediately from proposition 3.1, we have the values of Θ on the various regions of $\widetilde{PSL_2}\mathbb{R}$:

$$\Theta(\mathrm{Hyp}_n) = \left(\left(n - \frac{1}{2} \right) \pi, \left(n + \frac{1}{2} \right) \pi \right)$$

$$\Theta(\mathrm{Par}_n) = \left(\left(n - \frac{1}{2} \right) \pi, \left(n + \frac{1}{2} \right) \pi \right)$$

$$\Theta(\mathrm{Ell}_n) = \begin{cases} \left((n-1)\pi, n\pi \right) & \text{for } n > 0 \\ \left(-|n|\pi, (-|n|+1)\pi \right) & \text{for } n < 0 \end{cases}$$

From the inverse lemma 3.4, we immediately obtain $\Theta(\alpha) = -\Theta(\alpha^{-1})$. This also follows from the definition of Θ. We have $R(\alpha^{-1}) = R(\alpha)^{-1}$, and by continuity we can deduce $\tilde{R}(\tilde{\alpha}^{-1}) = \tilde{R}(\tilde{\alpha})^{-1}$, so that $\Theta(\tilde{\alpha}^{-1}) = -\Theta(\tilde{\alpha})$.

From proposition 3.7 we have that if $\mathcal{P}(\alpha, \beta; i)$ bounds an embedded (or immersed) disc, then

$$\Theta\left(\left[\tilde{\alpha}^{-1}, \tilde{\beta}^{-1} \right] \right) = \frac{1}{2} \Delta \left[\mathcal{P}\left(\alpha, \beta; i \right) \right]. \tag{4.1}$$

From lemma 3.9 we have:

$$2\Theta\left(\left[\tilde{\alpha}, \tilde{\beta} \right] \right) = \Delta[i, \beta i, \beta \alpha i] - \Delta[i, \alpha i, \alpha \beta i] - \Delta\left[i, \alpha \beta i, [\alpha, \beta] i \right], \quad \text{hence} \quad \left| \Theta\left(\left[\tilde{\alpha}, \tilde{\beta} \right] \right) \right| < \frac{3\pi}{2}. \tag{4.2}$$

The quasimorphism property of Θ follows immediately from the addition lemma 3.5.

$$\Theta\left(\tilde{\beta}\tilde{\alpha} \right) = \Theta\left(\tilde{\beta} \right) + \Theta\left(\tilde{\alpha} \right) - \frac{1}{2}\Delta[i, \beta i, \beta \alpha i], \quad \text{hence} \quad \left| \Theta\left(\tilde{\beta}\tilde{\alpha} \right) - \Theta\left(\tilde{\beta} \right) - \Theta\left(\tilde{\alpha} \right) \right| < \frac{\pi}{2} \tag{4.3}$$

Repeated application of the final inequality is enough to prove the Milnor–Wood inequality, as Milnor carries out in [18]. But the equalities involving areas in (4.1)–(4.3) appear to be new.

References

[1] Gérard Besson, Gilles Courtois, and Sylvestre Gallot, *Inégalités de Milnor-Wood géométriques*, Comment. Math. Helv. **82** (2007), no. 4, 753–803. MR MR2341839 (2009e:53055)

[2] Steven B. Bradlow, Oscar García-Prada, and Peter B. Gothen, *Maximal surface group representations in isometry groups of classical Hermitian symmetric spaces*, Geom. Dedicata **122** (2006), 185–213. MR MR2295550 (2008e:14013)

[3] Michelle Bucher and Tsachik Gelander, *Milnor-Wood inequalities for manifolds locally isometric to a product of hyperbolic planes*, C. R. Math. Acad. Sci. Paris **346** (2008), no. 11-12, 661–666. MR MR2423274 (2009f:53056)

[4] Marc Burger and Alessandra Iozzi, *Bounded differential forms, generalized Milnor-Wood inequality and an application to deformation rigidity*, Geom. Dedicata **125** (2007), 1–23. MR MR2322535 (2009c:53053)

[5] Peter Buser, *Geometry and spectra of compact Riemann surfaces*, Progress in Mathematics, vol. 106, Birkhäuser Boston Inc., Boston, MA, 1992. MR MR1183224 (93g:58149)

[6] Danny Calegari, *Circular groups, planar groups, and the Euler class*, Proceedings of the Casson Fest, Geom. Topol. Monogr., vol. 7, Geom. Topol. Publ., Coventry, 2004, pp. 431–491 (electronic). MR MR2172491 (2006i:57038)

[7] David Eisenbud, Ulrich Hirsch, and Walter Neumann, *Transverse foliations of Seifert bundles and self-homeomorphism of the circle*, Comment. Math. Helv. **56** (1981), no. 4, 638–660. MR MR656217 (83j:57016)

[8] Emmanuel Giroux, *Structures de contact sur les variétés fibrées en cercles audessus d'une surface (contact structures on manifolds that are circle-bundles over a surface)*, Comment. Math. Helv. **76** (2001), no. 2, 218–262. MR MR1839346 (2002c:53138)

[9] William M. Goldman, *Discontinuous groups and the euler class*, Ph.D. thesis, Berkeley, 1980.

[10] ———, *Topological components of spaces of representations*, Invent. Math. **93** (1988), no. 3, 557–607. MR MR952283 (89m:57001)

[11] ———, *The modular group action on real* SL(2)-*characters of a one-holed torus*, Geom. Topol. **7** (2003), 443–486 (electronic). MR MR2026539 (2004k:57001)

[12] Michael Gromov, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. (1982), no. 56, 5–99 (1983). MR MR686042 (84h:53053)

[13] Daniel Mathews, *From algebra to geometry: A hyperbolic odyssey; the construction of geometric cone-manifold structures with prescribed holonomy*, Masters thesis, University of Melbourne, 2005. Available at the author's website, `http://math.stanford.edu/~mathews`., 2005.

[14] ———, *Hyperbolic cone-manifold structures with prescribed holonomy I: punctured tori*, 2010.

[15] ———, *Hyperbolic cone-manifold structures with prescribed holonomy II: higher genus*, 2010.

[16] Dusa McDuff and Dietmar Salamon, *Introduction to symplectic topology*, second ed., Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1998. MR MR1698616 (2000g:53098)

[17] Geoffrey Mess, *Lorentz spacetimes of constant curvature*, Geom. Dedicata **126** (2007), 3–45. MR MR2328921 (2010a:53154)

[18] John Milnor, *On the existence of a connection with curvature zero*, Comment. Math. Helv. **32** (1958), 215–223. MR MR0095518 (20 #2020)

[19] Dennis Sullivan, *A generalization of Milnor's inequality concerning affine foliations and affine manifolds*, Comment. Math. Helv. **51** (1976), no. 2, 183–189. MR MR0418119 (54 #6163)

[20] John W. Wood, *Bundles with totally disconnected structure group*, Comment. Math. Helv. **46** (1971), 257–273. MR MR0293655 (45 #2732)